



**GDAŃSK UNIVERSITY  
OF TECHNOLOGY**

## **Podstawy kodowania mowy i obrazu**

Michał Hoefft

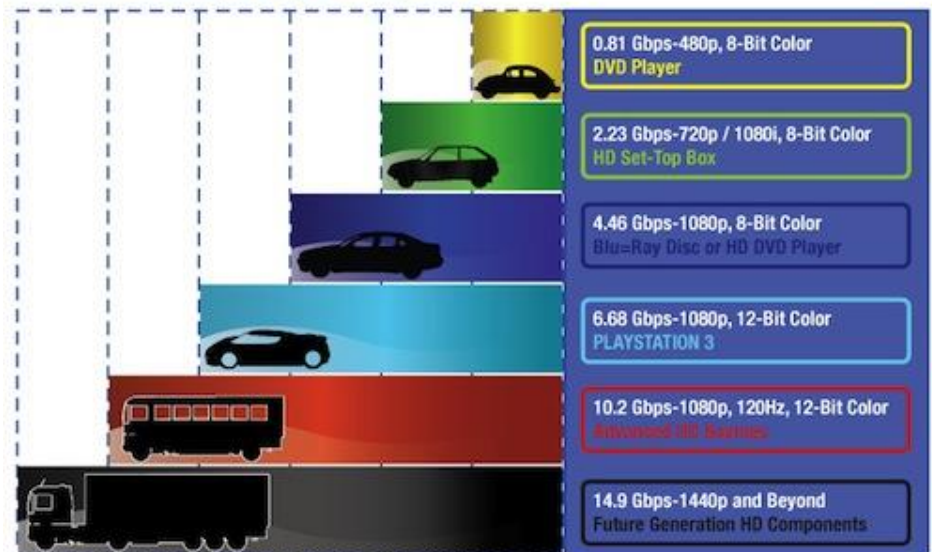
# Wstęp

- Kodowanie danych multimedialnych
  - Zapisanie w postaci cyfrowej danych multimedialnych pozwalające na ich późniejsze odtworzenie; przesłanie; archiwizację itp.

# Wstęp

- Ile zajmują dane multimedialne?
  - Standard telewizyjny (PAL 704x576; 12 bitów/piksel; 25 klatek/s)
    - Bez kompresji > 120Mbps  
prawie 110 GB  
dla filmu 120 min

HD Bandwidth & Applications  
HD Performance, 1440p and Beyond

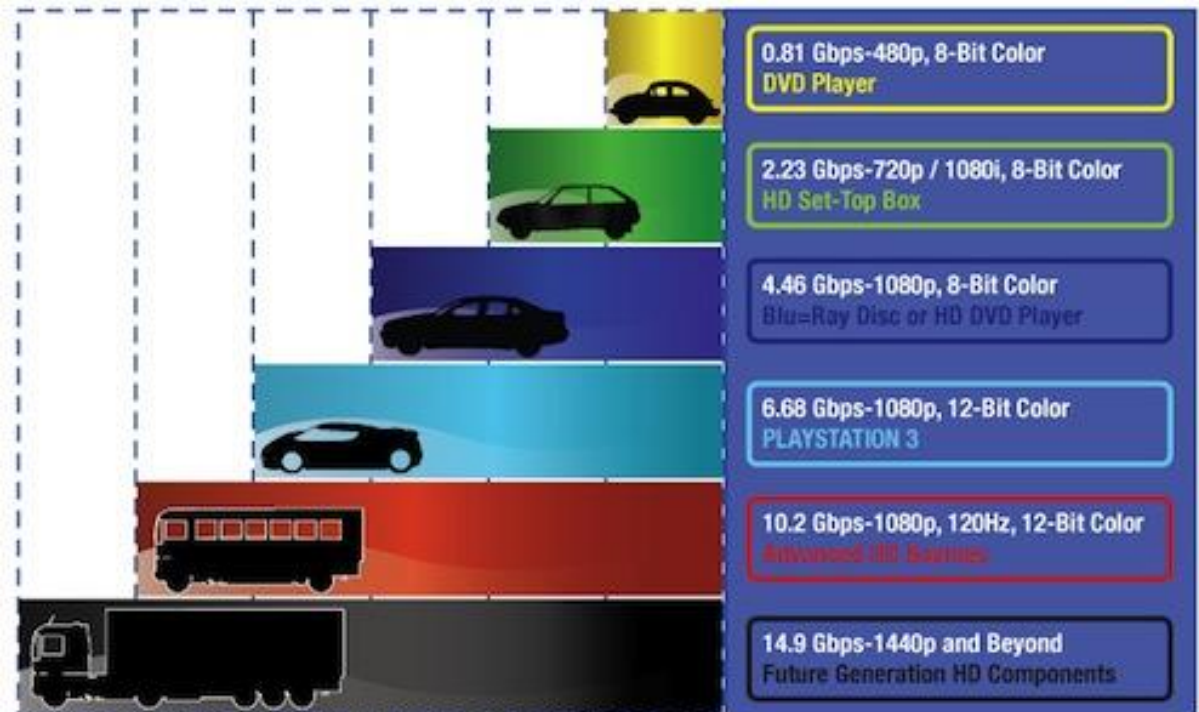


# Wstęp

- Ile zajmują dane multimedialne?

Youtube HD  
**6-8Mbps**  
Netflix HD  
**15Mbps**

**HD Bandwidth & Applications**  
HD Performance, 1440p and Beyond

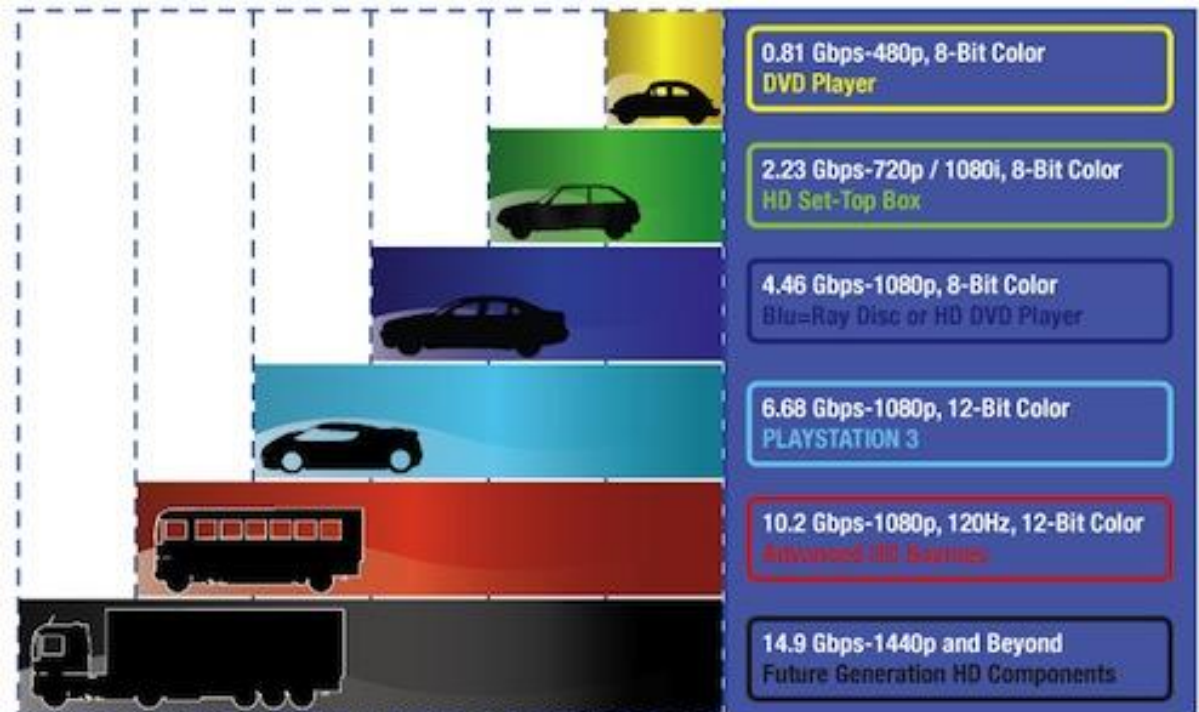


# Wstęp

- Ile zajmują dane multimedialne?

Youtube HD  
**6-8Mbps**  
Netflix HD  
**15Mbps**

**HD Bandwidth & Applications**  
HD Performance, 1440p and Beyond



# Wstęp

- Kompresja danych multimedialnych
  - Zmniejszenie ilości zasobów niezbędnych do przechowywania i przesyłania danych multimedialnych
  - Zwiększenie ilości zasobów niezbędnych do przetwarzania (odtworzenia) danych multimedialnych
- Kompresja stratna albo bezstratna

# Kompresja

- W procesie kompresji wykorzystujemy:
  - Występowanie nadmiarowości informacji
  - Zmiana sposobu reprezentacji (opisu) informacji
  - Ograniczenia perceptualne
    - Wzroku
    - Słuchu
- Metoda kompresji powinna być dostosowana do przetwarzanej informacji

# Ilość informacji

- Ilość informacji zawarta w wiadomości/zdarzeniu występującym z prawdopodobieństwem  $p$

$$I = \log_2 \left( \frac{1}{p} \right) [\text{bit}]$$

- Entropia – średnia ilość informacji w zbiorze wiadomości

$$E = \sum_i p_i I_i$$



# Ćwiczenie

- Jak zakodować symbole A, B, C i D jeżeli wiemy, że pojawiają się one odpowiednio z częstością 80%, 10%, 5% i 5%?

# Konwersja danych

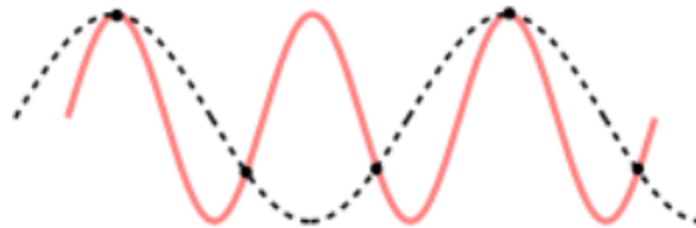
- Etapy konwersji:
  1. Przetwornik A/D
  2. Układ próbkujący
  3. Kwantyzer
  4. Koder

# Próbkowanie

- Jak często próbkować sygnał aby odzwierciedlić jego charakterystykę?

# Próbkowanie

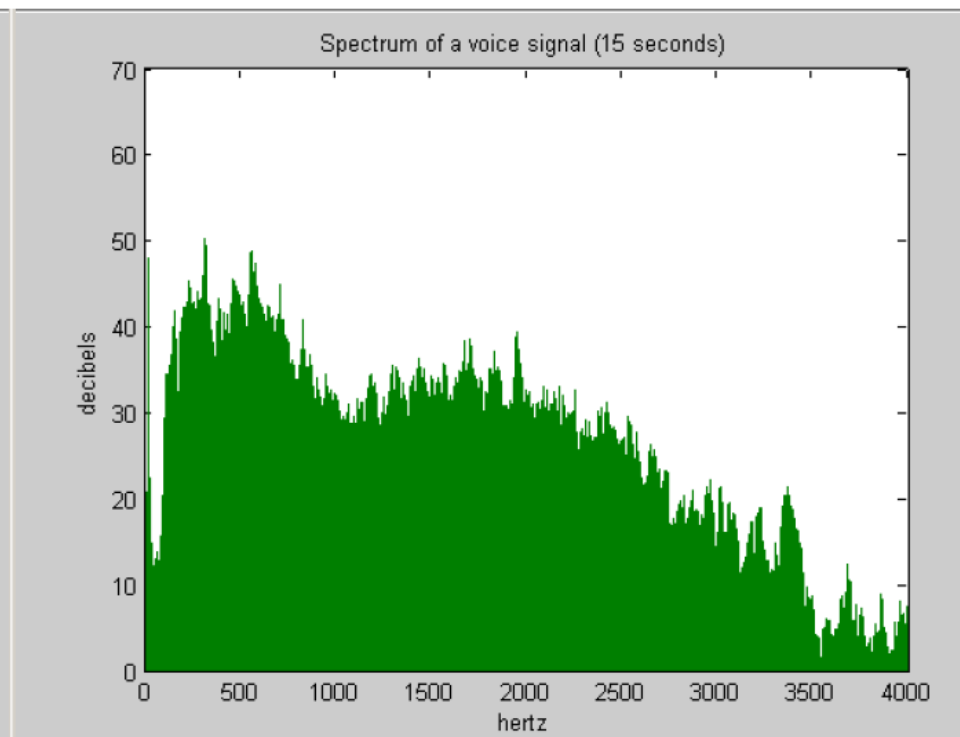
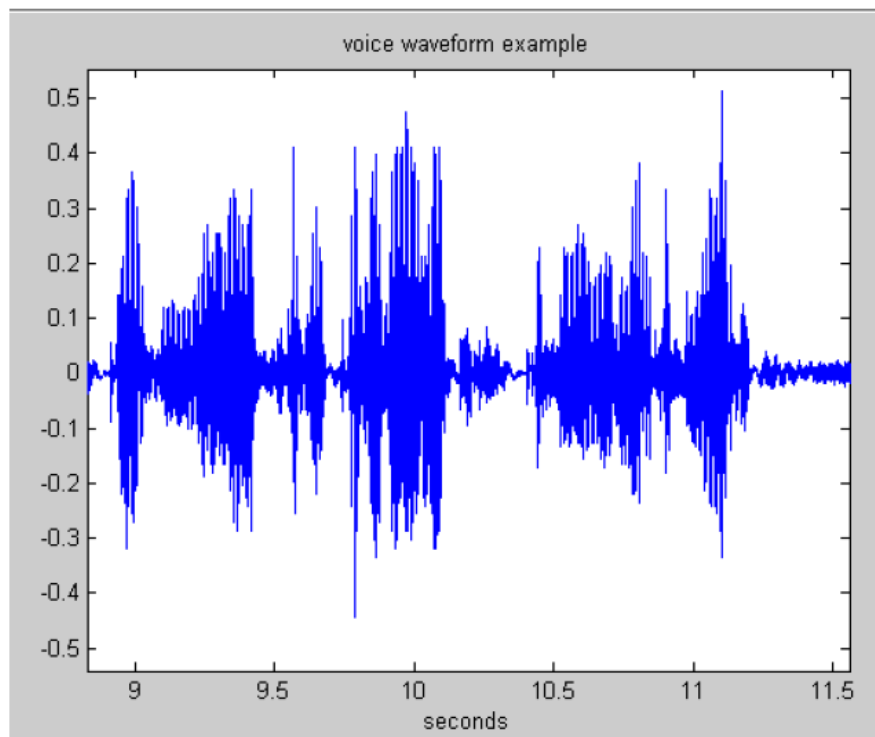
- Jak często próbkować sygnał aby odzwierciedlić jego charakterystykę?



## Twierdzenie o próbkowaniu

Jeśli sygnał ciągły nie posiada składowych widma o częstotliwości równej lub większej niż  $B$ , to może on zostać wiernie odtworzony z ciągu jego próbek tworzących sygnał dyskretny, o ile próbki te zostały pobrane w odstępach czasowych nie większych niż  $1/(2B)$ .

# Przebieg czasowy i widmo sygnału



- Często stosowane częstotliwości próbkowania

Częstotliwość	Zastosowanie
8.000	Telefonia
16.000	Telefonia szerokopasmowa
44.100	CD-Audio
96.000	DVD-Audio

# Kwantyzacja

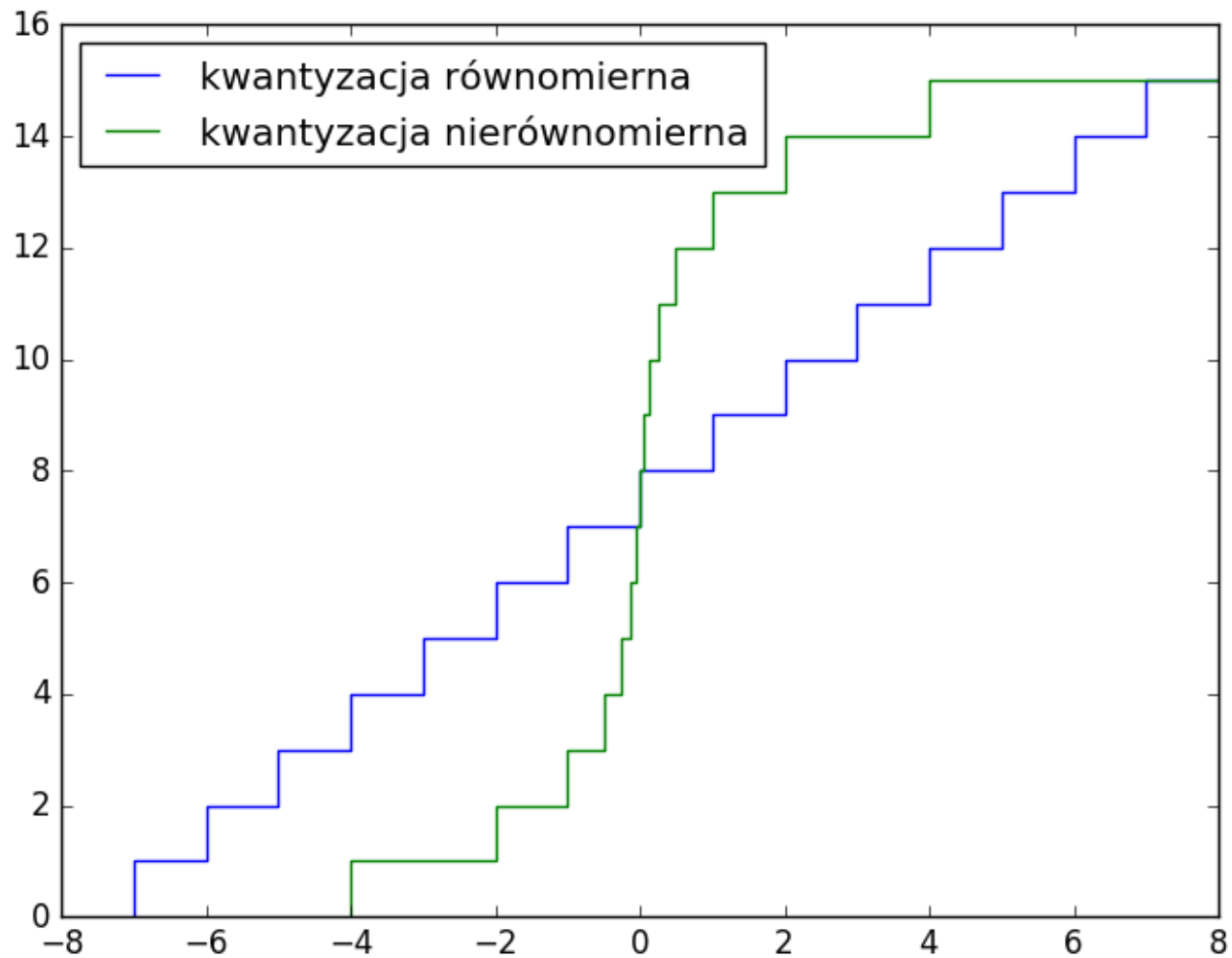
- Nieodwracalne odwzorowanie wartości próbki sygnału analogowego na wartość z ograniczonego zbioru wartości dyskretnych zmniejszające dokładność pomiaru wprowadzając szum kwantyzacji (błąd kwantyzacji)

# Kwantyzacja

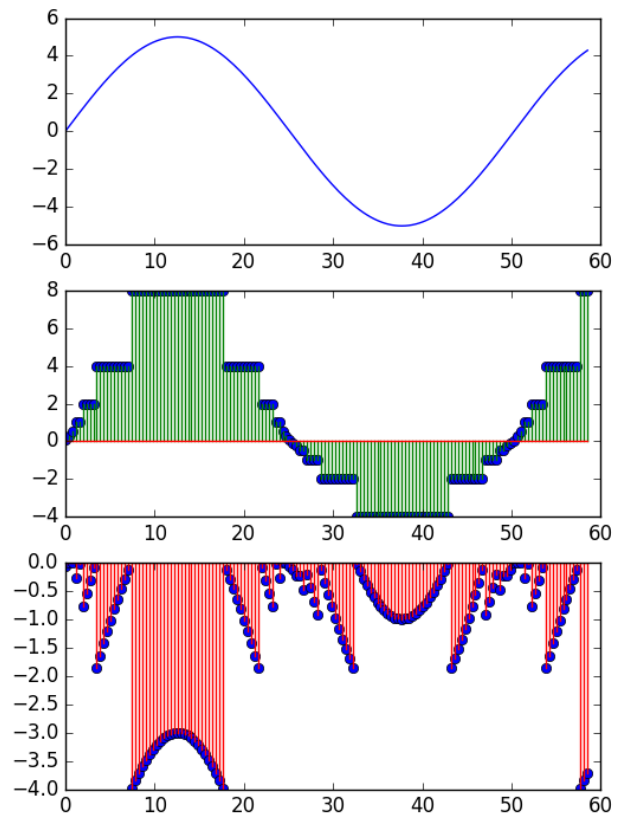
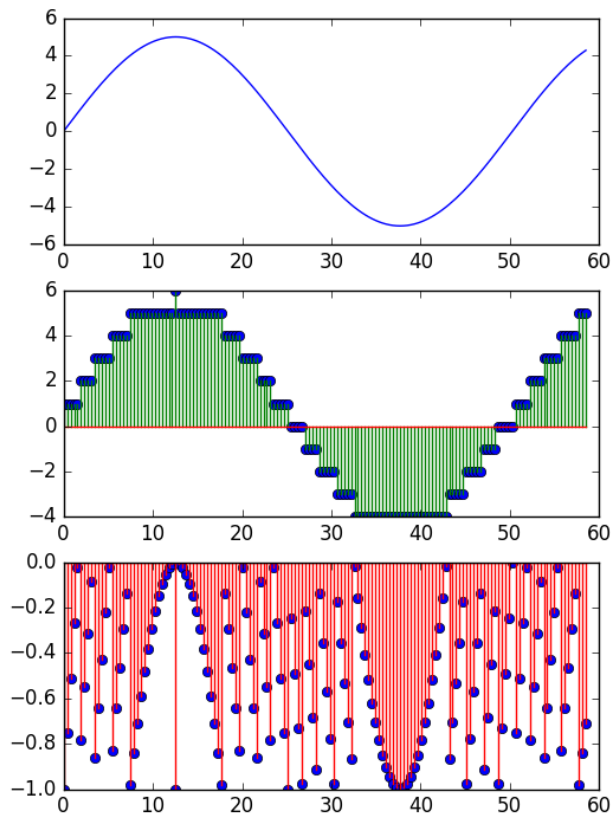
- Równomierna
  - Wszystkie przedziały jednakowe
  - Dobra dla sygnałów o jednorodnym prawdopodobieństwie wartości
- Nierównomierna
  - Przedziały o różnym rozmiarze
  - Dostosowanie do charakterystyki sygnału
  - Większa gęstość odwzorowanych wartości w obszarach większej zmienności sygnału



# Kwantyzacja



# DEMO

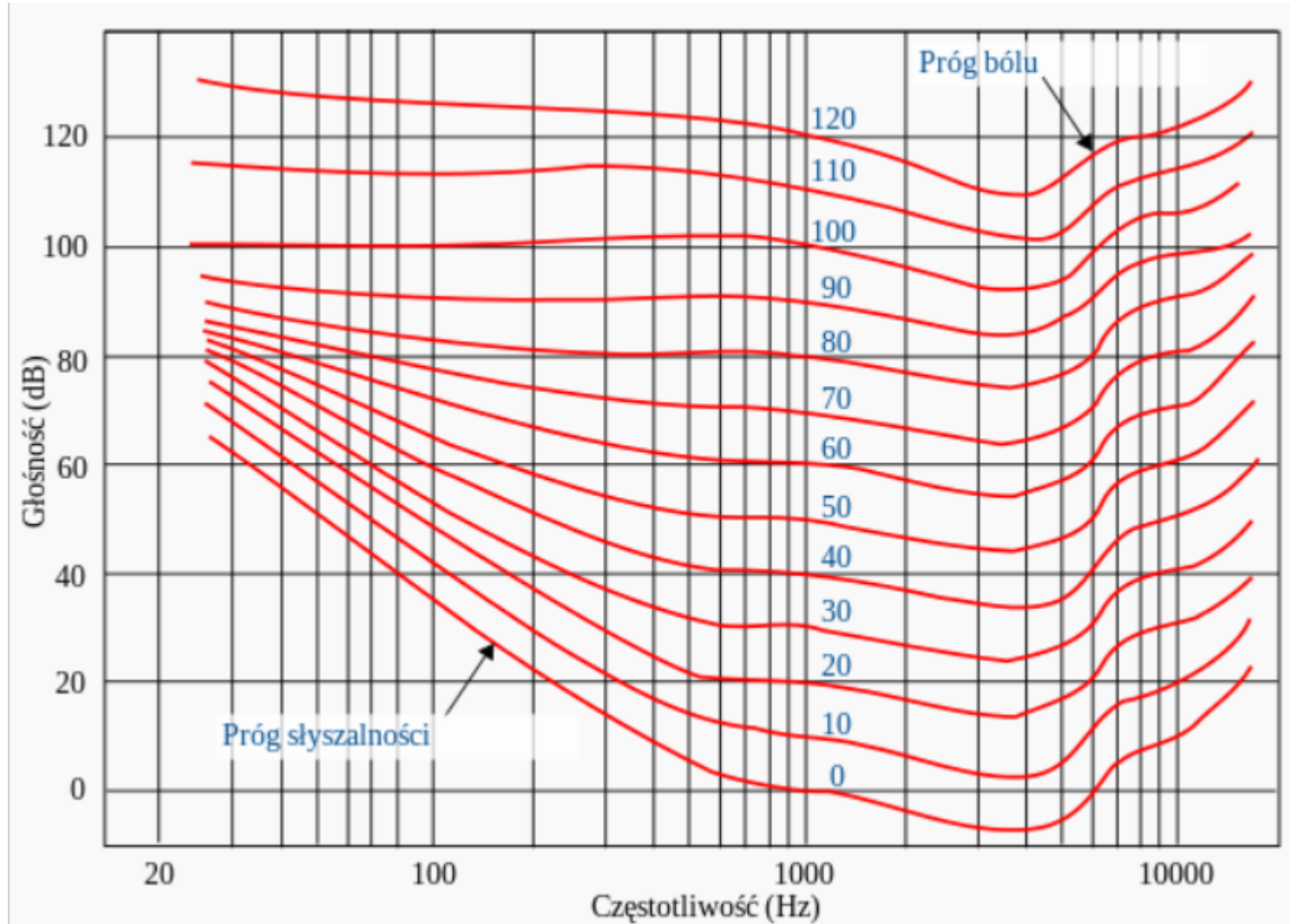


# Granice percepcji

- Częstotliwość
  - 16 Hz – 20 kHz
- Natężenie
  - 0 – 140 dB

Natężenie dźwięku [dB]	Okoliczności
130	Startujący odrzutowiec
100	Roboty uliczne (młot pneumatyczny)
80	Pociąg
40	Rozmowa
20	Biblioteka

# Krzywe izofoniczne



# Koder/Dekoder

- Zapisanie/Odczytanie ciągu wartości próbek w postaci sygnału
- Implementacja algorytmu kodowania/kompresji

# Kodowanie mowy

- Kodek G.711 PCM
  - Zastosowano kodowanie 14 bitowych próbek z kwantyzera równomiernego na 8 bitach logarytmicznie:
    - 1 bit znak (s)
    - 3 bity wykładnik (e)
    - 4 bity mantysa (m)

$$\text{ALaw } y = (-1)^s \cdot (16 \cdot \min\{e, 1\} + m + 0.5) \cdot 2^{\max\{e, 1\}}$$

- Zakres mowy do ok. 4000 Hz, częstotliwość próbkowania ?
- Strumień danych 64 kb/s

# Kodek G.711

- uLaw (USA, Japonia)

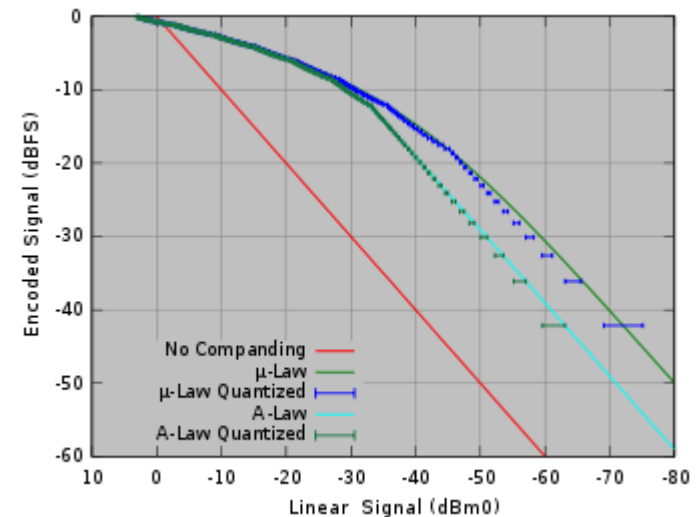
$$F(x) = \operatorname{sgn}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)} \quad -1 \leq x \leq 1$$

- ALaw (Europa, rozmowy międzynarodowe)

$$F(x) = \operatorname{sgn}(x) \begin{cases} \frac{A|x|}{1+\log(A)}, & |x| < \frac{1}{A} \\ \frac{1+\log A|x|}{1+\log(A)}, & \frac{1}{A} \leq |x| \leq 1, \end{cases}$$

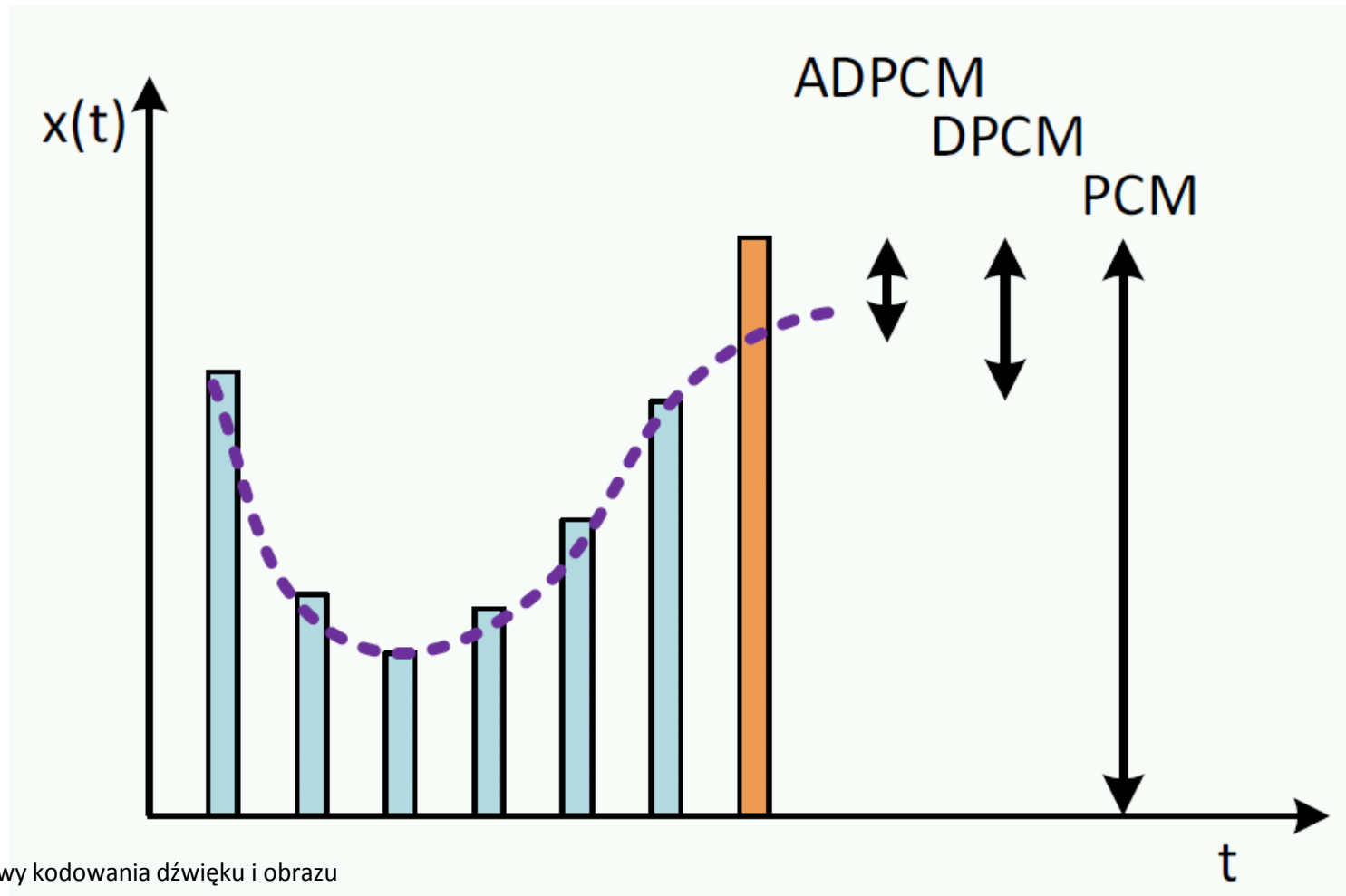
# ALaw vs uLaw

- uLaw
  - Zapewnia większą dynamikę przenoszonego sygnału
  - Kosztem większych zniekształceń dla sygnałów o małej amplitudzie

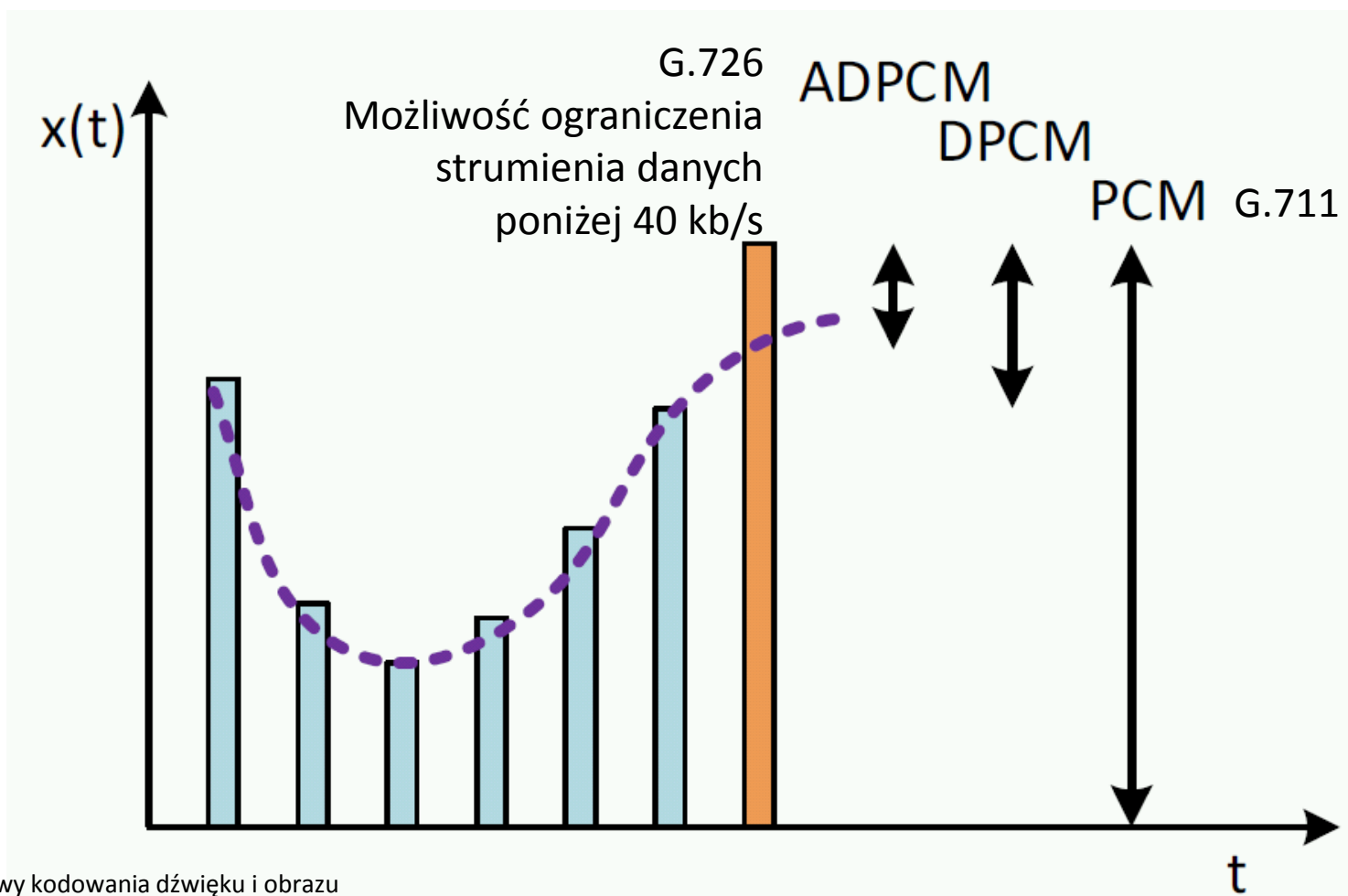




# Kodeki PCM



# Kodeki PCM



# Opóźnienie kompresji

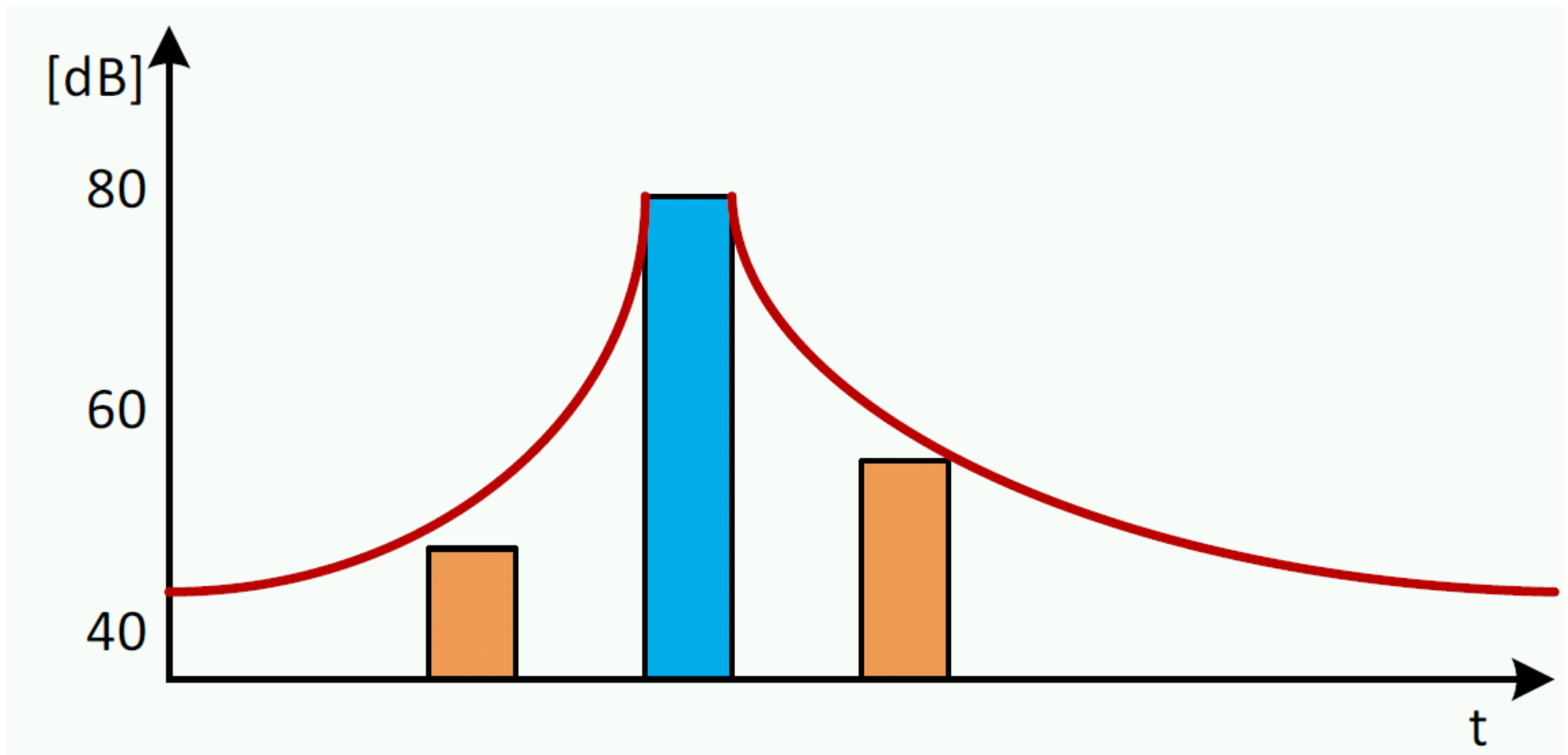
Na przykładzie urządzeń Cisco

Compression Method	Bit Rate (kbps)	MOS Score	Compression Delay (ms)
G.711 PCM	64	4.1	0.75
G.726 ADPCM	32	3.85	1
G.728 LD-CELP	16	3.61	3 to 5
G.729 CS-ACELP	8	3.92	10
G.729 x 2 Encodings	8	3.27	10
G.729 x 3 Encodings	8	2.68	10
G.729a CS-ACELP	8	3.7	10
G.723.1 MP-MLQ	6.3	3.9	30
G.723.1 ACELP	5.3	3.65	30

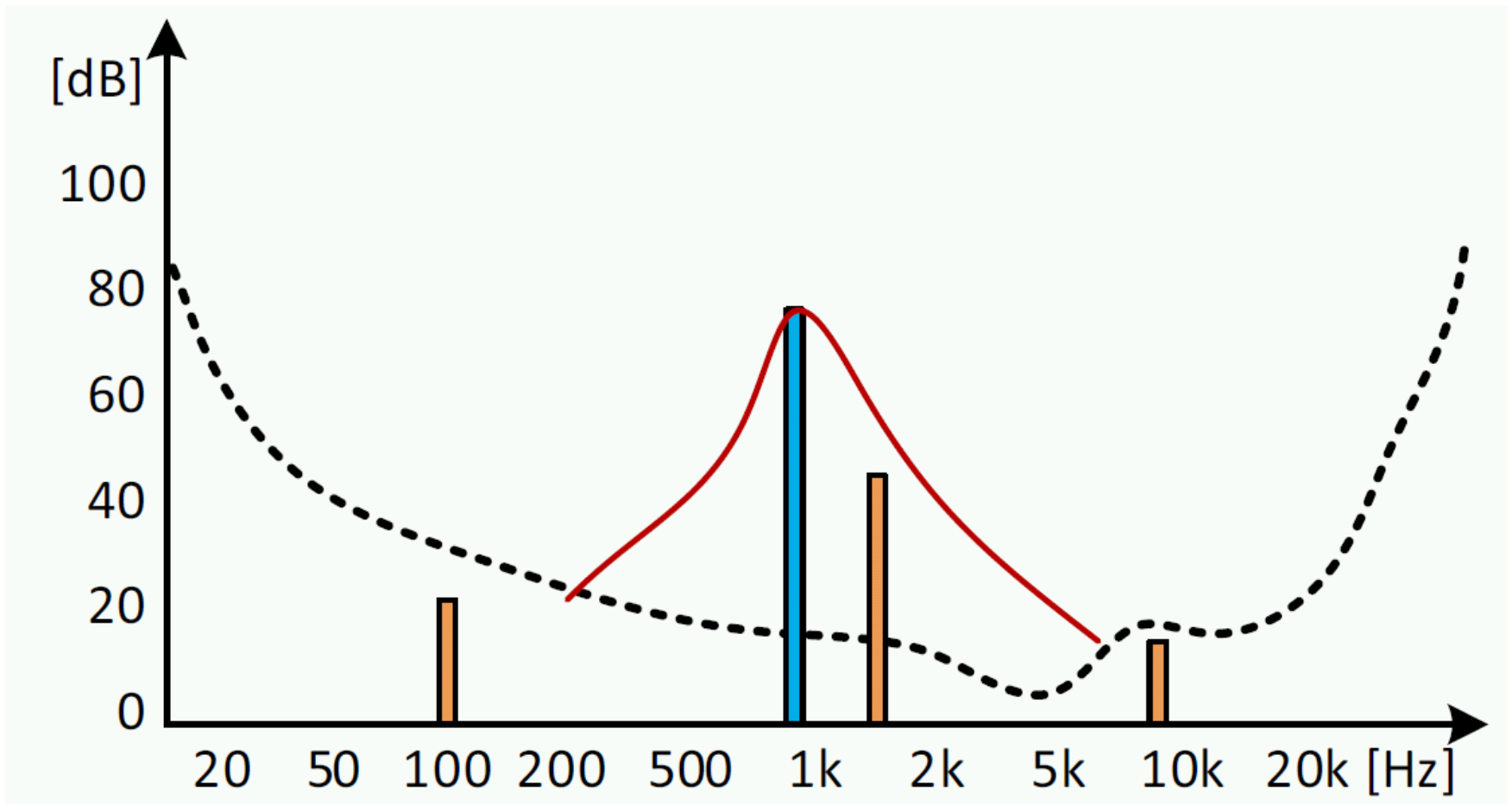
# Kodowanie dźwięku

- Odmienne wymagania niż w przypadku kodowania mowy
  - Inna dynamika
  - Inny zakres częstotliwości
- Większe możliwości stosowania maskowania perceptualnego
  - W dziedzinie czasu
  - W dziedzinie częstotliwości
  - Przykład MP3

# Maskowanie w dziedzinie czasu



# Maskowanie w dziedzinie częstotliwości



# Kodowanie obrazu

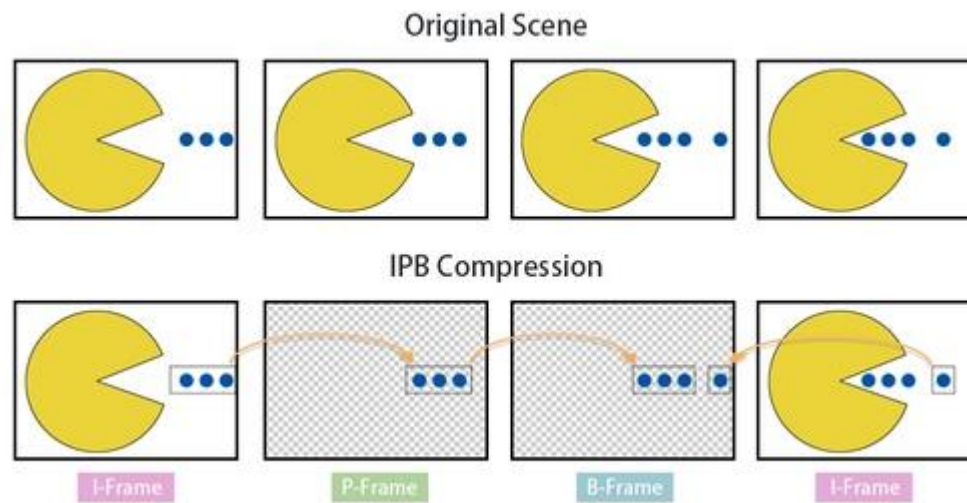
- Obraz nieruchomy JPEG:
  - Wykorzystanie modelu barw YCrCb
  - Zmniejszenie rozdzielczości składowych chrominalnych – ludzie oko jest mniej wrażliwe na niedokładności odwzorowania koloru niż jasności
    - 4:4:4 – nie zastosowano redukcji
    - 4:2:2 – redukcja zastosowana tylko w poziomie
    - 4:2:0 – redukcja zastosowana w poziomie i pionie
  - Cały obraz podzielony na obszary 8x8 pikseli
  - Dyskretna transformata kosinusowa (DCT)

# Kodowanie obrazu

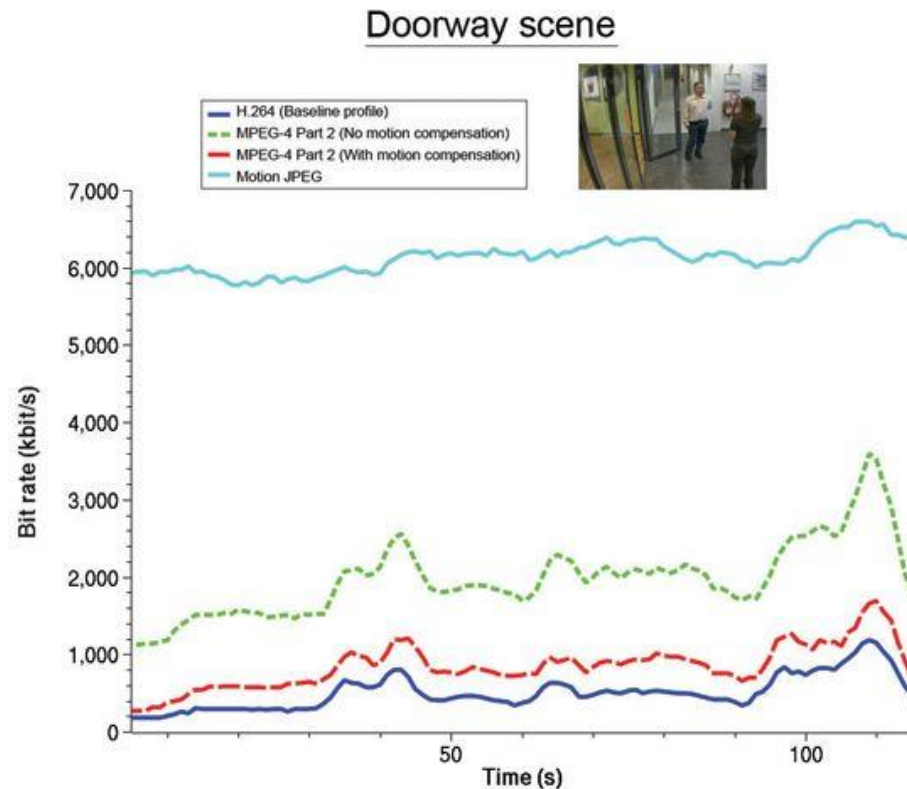
- Obraz ruchomy standard MPEG:
  - Bazuje na kodowaniu JPEG (wariant 4:2:0)
  - Wykorzystując fakt, że zwracamy większą uwagę na poruszające się elementy obrazu wprowadzono 3 rodzaje ramek:
    - I frame
    - P frame
    - B frame



- Ramki MPEG



- Korzyści z zastosowania kompresji



**DZIĘKUJĘ ZA UWAGĘ**